Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11455143 Little Holland Tract at North Breach near Courtland, California

This model archive summary summarizes the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC timeseries for the period of record October 16, 2014 o May 23, 2018. This is the first suspended-sediment model developed for the site. All data were collected using U.S. Geological Survey (USGS) protocols and are stored in the National Water Information System (NWIS) database. The methods used follow USGS guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.10 (USGS 2016) and USGS Techniques and Methods, book 3, chap C4 (Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS 2014).

Site and Model Information

Site number: 11455143
Site name: Little Holland Tract at North Breach near Courtland, California
Location: Latitude 38°20'07.48", Longitude 121°39'29.47" referenced to North American Datum of 1983, CA, Hydrologic Unit 18020163
Equipment: A YSI EXO2 sonde began logging turbidity on October 16, 2014.

Model number: 11455143.SSC.WY2015.1
Model calibration dataset period: November 23, 2015 to December 17, 2016.
Model application date: October 16, 2014 to May 23, 2018.
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Physical Sampling Details and Sediment Data

All sediment data were collected using U.S. Geological Survey (USGS) protocols (USGS, 2006) and are stored in the National Water Information System (NWIS) database https://waterdata.usgs.gov/nwis. Discrete, boat-based sample collection for SSC monitoring ideally occurs between 6-12 times per year. Sample collection spans the range of conditions and targets storm events during winter and spring flows as well as summer low flow conditions.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). Sediment samples represent the discharge-weighted concentrations of the stream cross section. The Equal Discharge Increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge weighted suspended-sediment samples were collected. Each sampling vertical is located at the centroid of increments representing 20% of the total flow (5 verticals). Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples to represent the average cross section because velocities are not always isokinetic (based on Table 4-5 from TWRI09A4, USGS

2006). A boat-based discharge measurement was collected immediately before sampling to determine the location of each vertical.

Technicians collected samples using a Federal Interagency Sediment Project (FISP) US D-74 depth-integrated suspended-sediment sampler. The channel cross-section across the breach is roughly 8.5 feet deep in the thalweg with a mean depth of approximately 6 feet. Sampling depths ranged from 4 to 8.3 feet depending on the tide and season. Station velocities typically ranged from -2.5 ft/sec during flood tides to +2.0 ft/sec during ebb tides. Sediment at this station is mostly fines (92% on average from sand/fine analysis) and potential sampling bias due to non-isokinetic sampling is considered minimal.

Samples were analyzed by the USGS Sediment Laboratory in Santa Cruz, California. All samples were analyzed for sediment concentration (mg/L) by the filtration method and most samples are also analyzed for the percentage of fines (< 0.062 mm). The sand/fine break analysis can be used to identify dataset variability and potential outliers and shows that sediment at this station is composed of mostly fines (92% fines on average). EDI verticals were analyzed individually by the lab. This method of individual analysis for quality control purposes because of rapidly changing, tidal conditions. The set average SSC of the five verticals was computed and used in the calibration model. In rare occasions when the SSC at a vertical was deemed an outlier, a manual average was computed from fewer than 5 verticals and occurred on November 23, 2015 and May 17, 2016 when the averages were computed from 4 verticals. Notes were applied to the database.

All sediment data were reviewed and marked as approved in the USGS NWIS Water-Quality System database (QWDATA) and made publicly available before being included in the calibration model. Publicly available field/lab sediment data can be found at: https://waterdata.usgs.gov/nwis/uv?site_no=11455143.

Surrogate Data

Continuous 15-minute turbidity data and discharge data were collected and computed by the USGS California Water Science Center and evaluated as possible explanatory variables for SSC. Data were measured using a YSI EXO2 sonde and reported in Formazin Nephelometric Turbidity Units (FNU). Data began logging on October 16, 2014 and the sonde was removed from the station on May 23, 2018. All surrogate turbidity data were computed, reviewed, and approved before using in the sediment calibration model per USGS guidelines (Wagner and others 2006). Discharge data were collected, computed, reviewed, and approved by the USGS California Water Science Center and retrieved from NWIS-TS. Methods to compute discharge follow Levesque and Oberg (2012). The 15-minute timeseries data are located at https://waterdata.usgs.gov/nwis/uv?site_no=11455143.

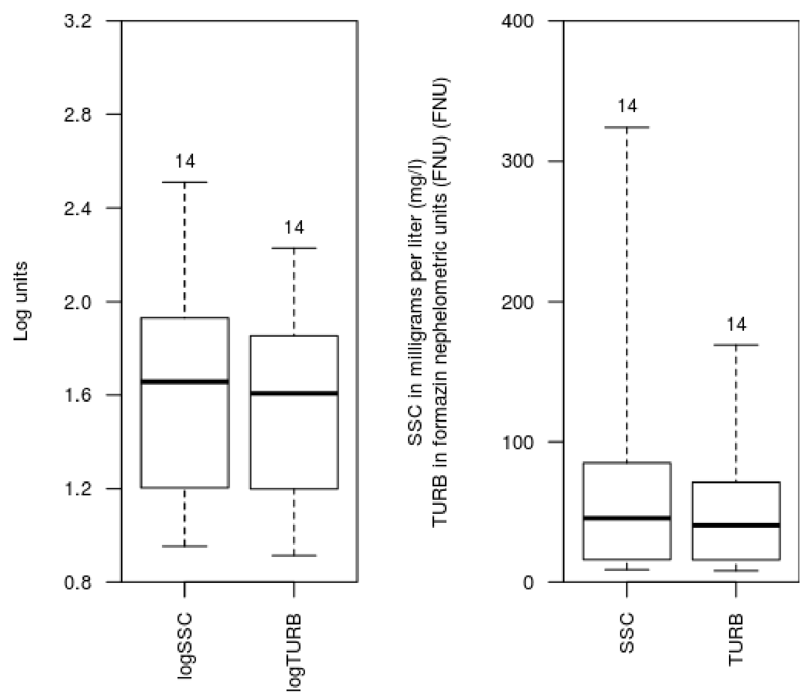
Model Calibration Dataset

The approved time-series turbidity data spanning the dates of the sediment constituent dataset were retrieved from NWIS-TS (Rasmussen and others 2009). The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate data with the discrete sediment data (Domanski and others 2015). Turbidity and discharge values were paired with each sediment sample observation from a matching max +/- of 15 minutes. The SAID manual is found at https://pubs.er.usgs.gov/publication/ofr20151177.

The regression model is based on 14 concurrent measurements of turbidity and SSC, and 3 concurrent discharge measurements. Summary statistics and the complete model-calibration data set are provided in the following sections.

Model Development

Multiple models were evaluated including simple linear regression (SLR) and multiple linear regression (MLR). The most common estimation technique is SLR, but MLR is an alternate tool for computing SSCs when the SLR model standard percentage error (*MSPE*) statistic is larger than 20 percent (Rasmussen and others, 2009). The calibration dataset is composed of 14 concurrent turbidity, SSC, and discharge measurements (discharge n=3). Boxplots are shown below for the turbidity and SSC data. USGS (2016) *recommends* a minimum of 36 paired observations, however the station was discontinued.



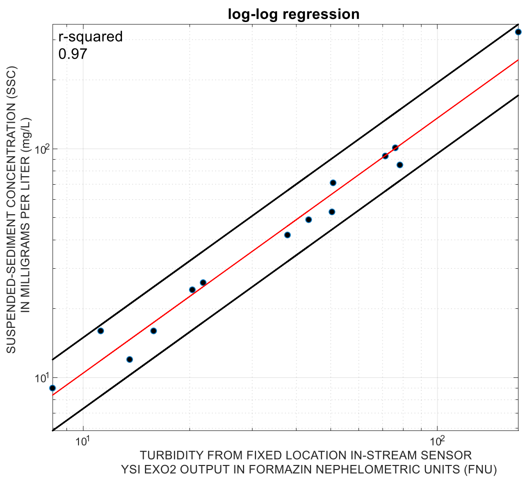
Model diagnostics and plots for model review were output using a variety of Matlab, SAID, and the R environment (R Core Team, 2018). The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate the models. The best model was chosen based on residual plots, model standard error, R^2, significance tests (p-values), correlation of explanatory variables, variance inflation factor (VIF), and PRESS (prediction error sum of squares) statistics. Values for the statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

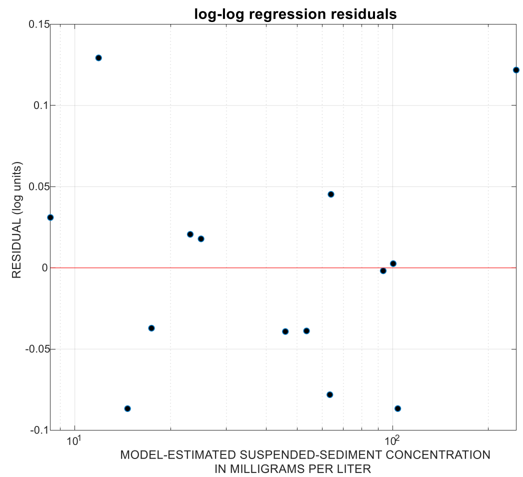
A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), and Model 5) $\log_{10}$ transformed model with two explanatory variables (turbidity and positive discharges). Diagnostic statistics are summarized below for the five models evaluated. Discharge was not considered further as a second surrogate (in addition to turbidity) for an MLR model because the discharge record was extremely limited. Note that the resulting R^2 and other diagnostics shown below for model 4 and 5 are not from robust MLRs.

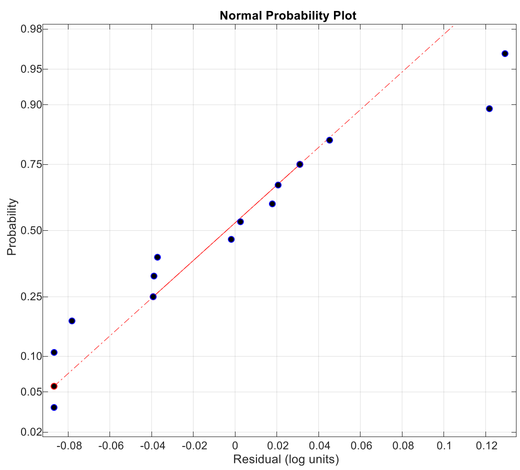
No.	R^2	R^2_a	RMSE	PRESS	MSPE	n	(type)
Model 1	0.95	0.94	19.2	17063	29.14	14	linear
Model 2	0.97	0.97	0.1	0.10	16.48	14	log
Model 3	0.84	0.83	33.4	139064	50.75	14	repeated median
Model 4	0.98	0.96	22.3	11407	33.95	3	multi-linear
Model 5	1.00	0.99	0.0	0.02	9.51	1	multi-log

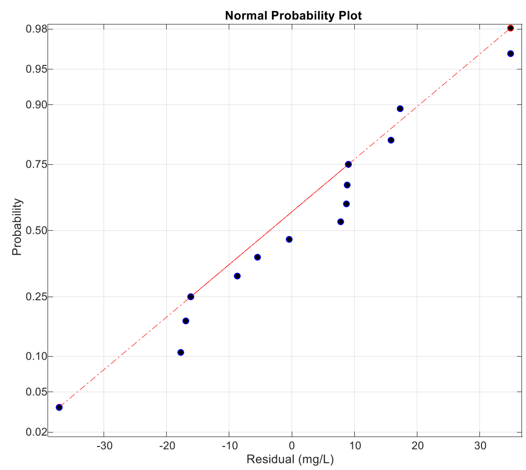
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to – 3 range are considered potential extreme outliers. The studentized residuals were reviewed from the SAID output reports and none of the samples were deemed as extreme outliers. All 14 observations were left in the model.

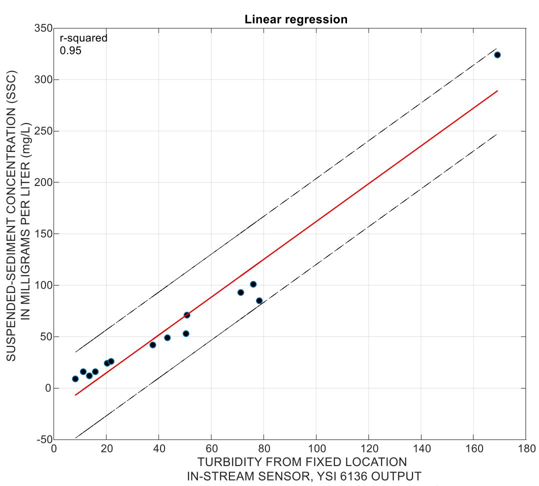
Of the SLR models, the $\log_{10}$-transformed model had the highest R^2, lowest error and the residuals plots for the $\log_{10}$-transformed regression (Model 2) indicates a more homoscedastic pattern (constant variance) and a more normal distribution compared to the linear model (see the graphs below comparing the log and linear SLRs). This model is significant, the model error (MSPE) was less than 20%, and the cross-validation also show good agreement amonst calibration data.

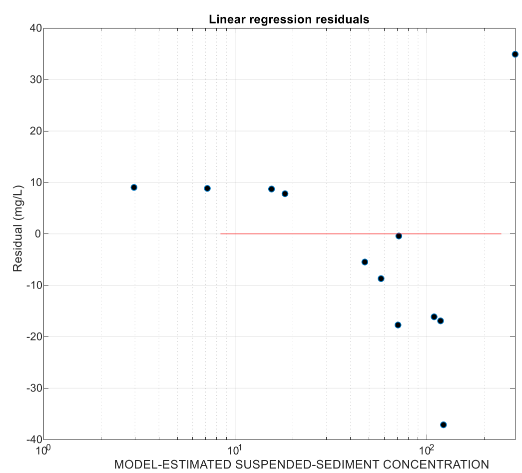

log-log regression
r-squared
0.97
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION IN-STREAM SENSOR
YSI EXO2 OUTPUT IN FORMAZIN NEPHELOMETRIC UNITS (FNU)


log-log regression residuals
RESIDUAL (log units)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER


Normal Probability Plot
Probability
Residual (log units)

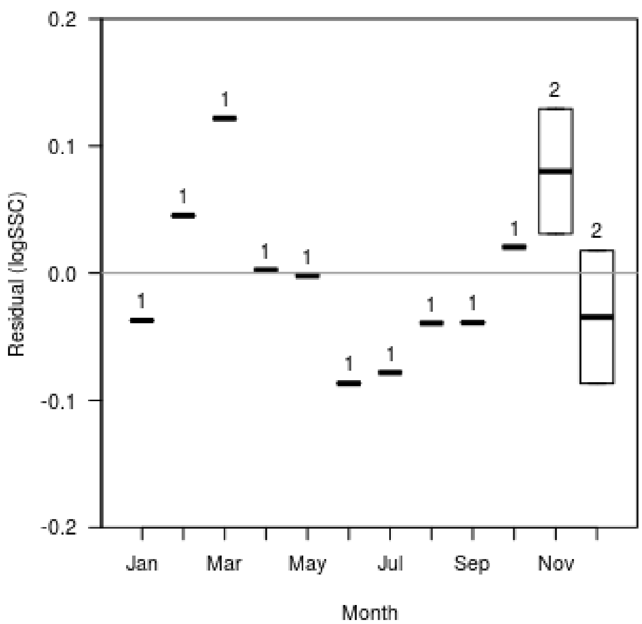

Normal Probability Plot
Probability
Residual (mg/L)

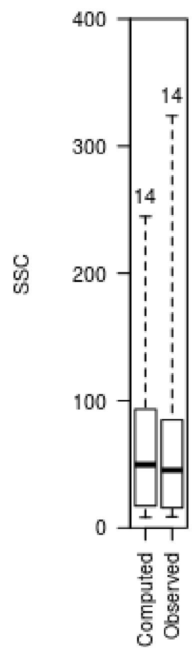

Linear regression
r-squared
0.95
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION
IN-STREAM SENSOR, YSI 6136 OUTPUT
IN FORMAZIN NEPHELOMETRIC UNITS (FNU)

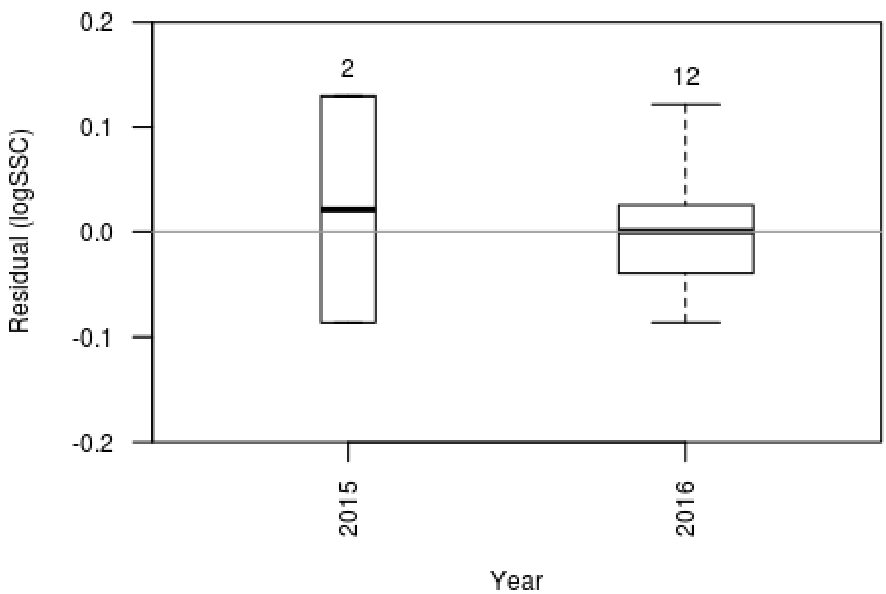

Linear regression residuals
Residual (mg/L)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER

Plots of log_{10}SSC and explanatory variables and residual diagnostic plots

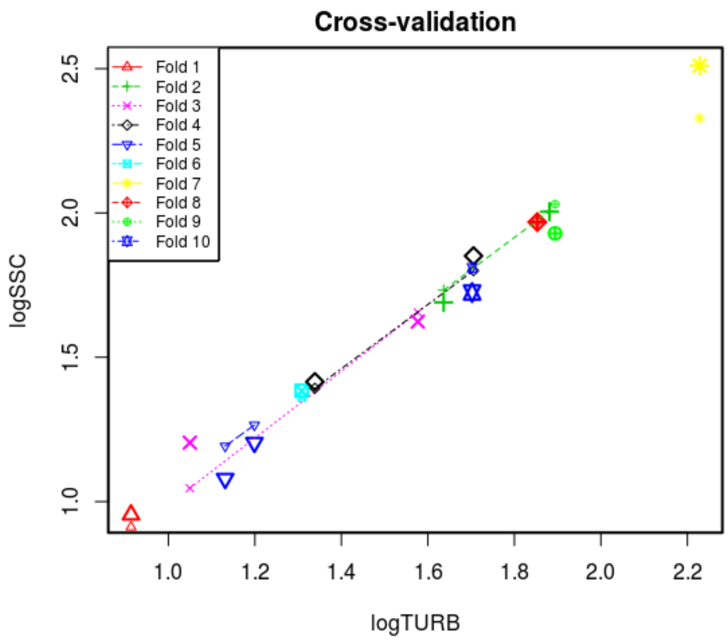
This summary is in accordance with the Office of Water Quality Technical Memorandum 2016.10 (USGS, 2016) which states this MAS **must** follow the format described in the memorandum. Based on this guidance, the following plots were generated using a specialized R-Script application specifically developed for this purpose by Patrick Eslick of the KSWSC (the MAS app) and is located at the following address:
https://patrickeslick.github.io/ModelArchiveSummary/



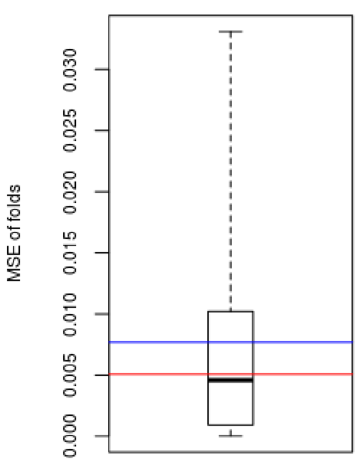




Cross-validation indicates when the model calibration data are randomly divided into subsets, the predictions from each subset regression model are very similar to the final regression model. The graph below shows a k-fold cross-validation with k=10 and the large points represent observations that were left out of each fold and are identified by the color and shape.




```
               Minimum MSE of folds:  4.32e-06
                  Mean MSE of folds:  7.67e-03
                Median MSE of folds:  4.57e-03
               Maximum MSE of folds:  3.31e-02
 (Mean MSE of folds) / (Model MSE):  1.51e+00
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

Definitions

```
SSC: Suspended sediment concentration (SSC) in mg/l (80154)
Turb: Turbidity in FNU (63680)
```

MAS App Version 1.0

Model Summary

The final regression model for suspended-sediment concentration for site 11455143 is a simple $\log_{10}$-transformed regression model based on 14 concurrent measurements of cross-sectional SSC and turbidity measurements. The simple linear regression model is shown below with basic model information, regression coefficients, correlation, summary statistics, and Duan's bias correction factor (Duan, 1983):

Linear Regression Model	**Coefficient of Determination (R^2)**
$log_{10}SSC = -0.0952 + 1.11 * log_{10}Turb$	0.97

$$\text{Log}_{10}(SSC) = -0.0952 + 1.11 \log_{10}(Turb)$$

where

SSC is suspended-sediment concentration, in milligrams per liter, and

Turb is turbidity, in formazin nephelometric units, measured with a YSI EXO2 sonde.

Because SSC was transformed during regression model development, the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) which is shown below.

Model	**Start date**	**End date**	**Linear Regression Model**	***BCF***
1	10/16/2014	5/23/2018	$SSC = 10^{-0.0952} \times Turb^{1.11} \times BCF$	1.01

The $\log_{10}$-transformed SLR model can be retransformed and corrected for bias resulting in the following equation:

$$SSC = 0.81 Turb^{1.11}$$

Parameter	Minimum	Maximum
Turbidity (FNU) entire record	3.6	566
Computed SSC (mg/L)	3.4	*921/360

*Extrapolation, defined as computation beyond the range of the model calibration dataset, may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model. The original maximum computed SSC beyond the extrapolation limit is 921 mg/L. The portion of time-series data beyond the extrapolation limit is less than 1%. Following USGS guidelines, the extrapolated, maximum computed SSC for this model is 360 mg/L.

Suspended-Sediment Concentration Record

The SSC record is computed using this regression model in the USGS National Real-Time Water Quality (NRTWQ) Web site. The complete record can be found at http://nrtwq.usgs.gov/ca.

Model

logSSC = + 1.11 * logTURB - 0.0952

Variable Summary Statistics

	logSSC	SSC	logTURB	TURB
Minimum	0.954	9.0	0.914	8.2
1st Quartile	1.200	16.0	1.200	15.8
Median	1.660	45.5	1.610	40.5
Mean	1.610	65.8	1.530	47.7
3rd Quartile	1.930	85.0	1.850	71.3
Maximum	2.510	324.0	2.230	169.0

Basic Model Statistics

Number of Observations	14
Standard error (RMSE)	0.0712
Average Model standard percentage error (MSPE)	16.5
Coefficient of determination (R^2)	0.975
Adjusted Coefficient of Determination (Adj. R^2)	0.973
Bias Correction Factor (BCF)	1.01

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	-0.0952	0.0816	-1.17	2.66e-01
logTURB	1.1100	0.0519	21.50	6.00e-11

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.972
E.vars	-0.972	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.429	0.192	0.756

Flagged Observations

Date	Time	logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
11/23/2015	11:34	1.20	1.07	0.129	2.02	2.38	0.194	0.491	1.17
3/15/2016	11:19	2.51	2.39	0.122	2.09	2.51	0.330	1.080	1.76

Model-Calibration Data Set

Observation Number	DateTime	logSSC	logTURB	SSC	TURB	Computed logSSC	Computed SSC	Residual	Normal Quantile	Censored Values
1	11/23/2015 11:34	1.20	1.05	16	11.2	1.07	12	0.129	1.725	--
2	12/17/2015 9:50	1.08	1.13	12	13.5	1.17	15	-0.087	-1.725	--
3	1/11/2016 11:28	1.20	1.20	16	15.8	1.24	18	-0.037	-0.268	--
4	2/19/2016 10:34	1.85	1.71	71	50.8	1.81	65	0.045	0.904	--
5	3/15/2016 11:19	2.51	2.23	324	169	2.39	248	0.122	1.212	--
6	4/27/2016 10:30	2.00	1.88	101	76.1	2.00	102	0.003	0.088	--
7	5/17/2016 11:08	1.97	1.85	93	71.3	1.97	95	-0.002	-0.088	--
8	6/8/2016 8:56	1.93	1.89	85	78.3	2.02	105	-0.087	-1.212	--
9	7/11/2016 10:32	1.72	1.70	53	50.4	1.80	64	-0.078	-0.904	--
10	8/9/2016 10:55	1.62	1.58	42	37.7	1.66	47	-0.039	-0.663	--
11	9/20/2016 12:03	1.69	1.64	49	43.3	1.73	54	-0.039	-0.457	--
12	10/20/2016 12:55	1.38	1.31	24	20.3	1.36	23	0.021	0.457	--
13	11/17/2016 12:21	0.95	0.91	9	8.2	0.92	8	0.031	0.663	--
14	12/17/2016 11:23	1.41	1.34	26	21.8	1.40	25	0.018	0.268	--

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., http://doi.org/10.3133/20151177.

Duan, Naihua. 1983. Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association. Volume 78-383. 605-610 p.

Edwards TK and Glysson GD. 1999. Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 3, Chap. C2. 89 p. Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. (Also available at http://pubs.usgs.gov/tm/3a23/.)

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: https://www.R-project.org/.

Rasmussen P, Gray JR, Glysson GD, Ziegler AC. 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data. Book 3 Applications of Hydraulics, Section C. 52 p. Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey. 2006. National field manual for the collection of water quality data: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 9, Chapter A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

[USGS] U.S. Geological Survey. 2016. Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07. Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.